PERGAMON

Molecular Immunology 38 (2001) 289–302

Molecular Immunology

www.elsevier.com/locate/molimm

Germ-line transcripts of the immunoglobulin λ J–C clusters in the mouse: characterization of the initiation sites and regulatory elements

Holger Engel [a,1], Holger Rühl [b], Craig J. Benham [c], Jürgen Bode [b], Siegfried Weiss [a,*]

[a] *Department of Cellbiology and Immunobiology, GBF, German Research Centre for Biotechnology, Mascheroder Weg 1, 38124 Braunschweig, Germany*

[b] *Department of Gene Regulation and Differentiation, GBF, German Research Centre for Biotechnology, Mascheroder Weg 1, 38124 Braunschweig, Germany*

[c] *Department of Biomathematical Science, Mount Sinai School of Medicine, New York, NY 10029, USA*

Received 26 April 2001; accepted 12 June 2001

Abstract

Transcription of unrearranged immunoglobulin gene segments strongly correlates with their accessibility to the V(D)J recombination machinery. The regulatory mechanisms governing this germ-line transcription are still poorly defined. In order to identify new regulatory elements, we first carried out a detailed characterization of the transcription initiation sites for the J–C germ-line transcripts, using rapid amplification of 5′ cDNA ends, assisted by a template switching mechanism at the 5′-end of the RNA. Transcripts were observed that initiated heterogeneously, starting up to 293 (λ1), 116 bp (λ2) and 79 bp (λ3) upstream from the respective Jλ gene segment. Additional RT-PCR analysis revealed the existence of germ-line transcripts of λ and also of κ that initiate even more upstream of these transcription initiation sites, although their frequencies were low. Promoter activity was detected in vitro 5′ of Jλ2, with the minimal promoter activity mapping to the region between positions −35 and −120. In addition, computer analysis allowed the prediction of a nuclear scaffold/matrix attachment (S/MAR) region between the two J–C gene clusters at each hemi-locus. This region between the λ1/λ3 clusters binds to the nuclear matrix in vitro, and J–C λ1 germ-line transcription initiates a short distance downstream from this S/MAR element.

Keywords: B lymphocytes; Transcription; Gene rearrangement; Gene regulation; mRNA

1. Introduction

The antigen receptors of B and T cells are assembled from gene segments that are physically separated in the germ-line genome. A hierarchy of events exists due to which, for example, the sequential rearrangements of first DJ and then VDJ at the IgH chain locus take place prior to the assembly of V and J at the IgL chain loci. In the mouse, the κ gene segments also start to rearrange at an earlier developmental stage than the gene segments of λ (Engel et al., 1999, for review see Gorman and Alt, 1998).

Strong evidence exists that both rearrangement at a particular locus and Ig class switch recombination, are preceded by transcriptional activity of the respective germ-line gene segments (Schlissel and Baltimore, 1989; Lennon and Perry, 1990; Engel et al., 1999). It remains unclear, however, whether the appearance of germ-line transcripts either requires or induces an alteration in the accessibility of the locus. However, promoter activity at germ-line loci seems to be essential for efficient rearrangement. Deletion of the promoter regions of both germ-line Cκ transcripts led to a dramatic reduction in the frequency of rearrangement at the targeted κ locus (Liu and van Ness, 1999; Cocea et al., 1999). A comparable effect was also found at the TCR loci: deletion of the germ-line promoter upstream from Dβ1 led to diminished rearrangement of this gene segment

* Corresponding author. Tel.: +49-531-6181-230; fax: +49-531-6181-444.

E-mail address: siw@gbf.de (S. Weiss).

[1] Present address: Qiagen GmbH, Max-Volmer-Str. 4, 40724 Hilden, Germany.

0161-5890/01/$ - see front matter
PII: S0161-5890(01)00056-6

(Whitehurst et al., 1999). Similarly, usage of those J segments in the TCR α locus that are proximal to the V gene segments was drastically reduced when the initiation region of the Jα germ-line transcript (TEA) was deleted (Villey et al., 1996).

To date, only the germ-line transcripts of murine Cμ and Cκ have been studied in detail, although germ-line transcripts from other Ig gene segments have been described (Gorman and Alt, 1998). At the germ-line μH chain locus, Iμ transcription initiates at heterogeneous start sites within the 5′ region of the Eμ enhancer located in the intron between J_H-Cμ and proceeds both through that intron and through Cμ. The processed form of this transcript thus contains the I exon derived from that intron, spliced to Cμ (Lennon and Perry, 1985). A second transcript, μ^0, initiates upstream of D_{Q52} and contains in its processed form the whole region from D_{Q52} down to the J_H1 gene segment, spliced to the Cμ region (Alessandrini and Desiderio, 1991).

Two germ-line transcripts also have been described for the murine Cκ region. A primary transcript of ≈8.4 kb initiates 3.5 kb upstream from Jκ1 and is processed into a 1.1 kb mRNA. A second transcript 4.7 kb in length initiates just upstream of Jκ1 and is processed into a 0.8 kb mRNA (Van Ness et al., 1981; Leclercq et al., 1989). The low abundance of these murine germ-line transcripts that are observed and the presence of multiple start/stop sites in all three reading frames suggest that they may not encode proteins. For this reason, they are sometimes called 'sterile' transcripts. However, a 15 kD protein potentially could be produced from the human germ-line J–C κ transcript by use of a GUG start codon (Francés et al., 1994).

Little is known about the elements that regulate the differential appearance of germ-line transcripts from the various Ig loci. The promoter of the Iμ transcript has been found to be located inside and possibly to be identical with the Eμ enhancer (Su and Kadesch, 1990). Transcription from this promoter appears to be under the control of the *E2A* gene product, since expression of Iμ was diminished in pro B cells of *E2A*-deficient mice (Bain et al., 1994). An octamer motif has been identified 5′ to the initiation sites of both Cκ germ-line transcripts (Martin and Van Ness, 1990). However, transcription could also be activated in vitro from one of these promoters in the absence of the octamer motif (Prabhu et al., 1996).

To date, almost no investigations have been made on the biology of germ-line transcripts derived from the murine λ L chain locus. This group has started to characterize these transcripts in the pro B cell line *R2-bfl*. This line can be grown as undifferentiated pro-B cells on bone marrow stromal cells in the presence of IL-7. Upon removal of IL-7, *R2-bfl* cells differentiate to the phenotype of mature B cells, with the capacity to switch their IgH chain isotypes (Rolink et al., 1996). However, due to their RAG-2 deficiency, these cells cannot rearrange their Ig loci. Thus, all Ig transcripts are products of gene segments in the germ-line configuration.

R2-bfl cells have previously been used to study the induction of transcripts containing the C regions from either the μ H, κ L or λ L chain germ-line gene clusters (Grawunder et al., 1995b; Engel et al., 1999). Interestingly, upon differentiation of *R2-bfl* cells the J–C λ germ-line transcripts appeared 2 days after the germ-line κ transcripts, which was also confirmed using pre-B cells isolated from bone marrow of mice (Engel et al., 1999). These data suggest that the consecutive activation of these two L chain loci – κ first, then λ – could result in a significant preference for functional rearrangements at the κ L chain locus and thus contribute to the extreme preponderance of κ L chains in the mouse.

In this first attempt to characterize the regulatory elements involved in differential activation of the murine L chain loci, we investigate the initiation sites of J–C λ germ-line transcripts using rapid amplification of 5′ cDNA ends, assisted by a template switching mechanism at the 5′ end of the RNA (SMART-5′-RACE) (Chenchik et al., 1998). The results obtained are validated by determining the initiation sites of the two germ-line transcripts from the μ H chain and the κ L chain locus, respectively, using the same approach. Further validation is achieved by using this technique to determine the initiation sites of mb-1 (Igα), a well-characterized gene with moderate heterogeneity of transcriptional initiation. In addition, we identified regulatory regions such as promoter and scaffold/nuclear matrix attachment regions (S/MARs) upstream from the initiation sites of the J–C λ transcripts. S/MARs represent sites within eukaryotic chromosomes where the DNA attaches to the nuclear scaffold (also called the nuclear matrix). S/MARs are thought to have important modulating effects on transcriptional regulatory events in the eukaryotic genome. We show that transcriptional initiation of the germ-line J–C λ transcripts occurs just downstream from sites that have strong S/MAR potential, suggesting that scaffold binding could play a role in their regulation.

2. Materials and methods

2.1. Cell lines and culture conditions

The B lineage cell lines *70Z/3* (Paige et al., 1978), *A20* (Kim et al., 1979), *J558* (Weigert et al., 1970), *S194* (Horibata and Harris, 1970), P3X63Ag8.653 (*X63/0*) (Kearney et al., 1979) and the thymoma cell line *BW 5147* (Ralph, 1973) were cultured in Optimem-1

medium (Gibco, Eggenstein, Germany) containing 5% FCS, 10 μg/ml penicillin/streptomycin and 50 nM β-mercaptoethanol. The rearrangement-deficient, bcl-2 transgenic pro B cell line *R2-bfl* was grown on stromal cells in the presence of recombinant IL-7, as described previously (Grawunder et al., 1995b). Differentiation was achieved by removing IL-7 and culturing on stromal cells for an additional 5 days. These cells are termed *R2-bfl* d5 cells.

2.2. SMART-5′-RACE of μ, κ, λ and mb-1

Total RNA from *R2-bfl* d5 and *X63/0* cells was isolated with Trizol reagent (Gibco) following the protocol supplied by the vendor. A total of 2 μg of RNA was used to generate 5′-RACE cDNAs using the SMART-RACE cDNA amplification kit (Clontech, Heidelberg, Germany) and the supplied protocols. These cDNAs were diluted to 100 μl with the tricine-EDTA buffer contained in the kit. The following primers were used to amplify 5′ ends. For λ germ-line transcripts: Cλ1-479; Cλ1-285; Cλ1-R2; Cλ2RE2; Cλ3RE2, all described in Engel et al. (1999); κ0.8 and κ1.1: Cκrev (Grawunder et al., 1995a); Iμ and μ^0: Cμrev 5′-CATCTGAACCTTCAAGGATGCTC-3′. mb-1rev: 5′-TCATGGCTTTTCCAGCTGGGC-3′; mb-1rev2: 5′-GAGGTTGCCCACATCCTGGTAGG-3′. PCR reactions were carried out in 50 μl volumes containing 200 μM of each dNTP, 200 nM of the gene-specific primer, a 1 × concentration of the universal primer mix supplied with the SMART-RACE kit, 2.5 μl of cDNA, 2.0 U Hotstart Taq DNA Polymerase (Qiagen, Hilden, Germany) and 5.0 μl of the 10 × PCR buffer supplied with the Taq polymerase. Each PCR program was initiated by activation of the Hotstart Taq polymerase at 95 °C for 15 min. The cycle conditions used were first 20 s at 95 °C, followed by 15 s at 68 °C, then 2 min at 72 °C. After 15 cycles the annealing temperature was decreased to 65 °C for ten cycles, followed by an additional ten to 20 cycles using 62 °C for the annealing temperature, the exact number depending on the efficiency of each RACE-PCR reaction. The presence of products of the expected sizes was determined by analyzing 5 μl from each reaction after a total of either 30, 35 or 40 cycles on a 2 or 3% agarose gel stained with ethidium bromide.

PCR products were gel purified and cloned using the TOPO TA cloning kit (Invitrogen, Leek, Netherlands). Random clones were picked and expanded and plasmid DNA was extracted. DNA was sequenced by cycle sequencing and analysis on an ABI 377 sequencer (Applied Biosystems, Heidelberg, Germany) with either the M13 forward or reverse primer supplied with the TOPO TA cloning kit. Sequences were analyzed using Sequencher software (Gene Codes Corp., Ann Arbor, MI).

2.3. RT-PCR of κ and λ germ-line transcripts

Total RNAs of BALB/c spleen cells and *R2-bfl* d5 cells were isolated as described above and 1 μg was reverse transcribed in a total volume of 20 μl using oligo-dT primers and Superscript reverse transcriptase, following the supplied protocols. After cDNA synthesis, 80 μl of water were added and 1 μl of this dilution was used as a template for RT-PCR. Water (1 μl) was added as a negative control. All PCRs were performed in a 50 μl reaction volume containing 1.25 U AmpliTaq Gold, 1 × reaction buffer, 200 μM of each dNTP and 200 nM of each primer set. After an initial preincubation to activate the polymerase for 10 min at 95 °C, the following cycle parameters were used: 20 s at 95 °C, followed by 1 min at 58 °C, then 2 min at 68 °C. Some 10 μl of each reaction was removed after 27, 30 and 33 cycles (27, 31 and 35 for λ2), and analyzed on a 2% agarose gel stained with ethidium bromide. The forward primers used were (the numbers given in brackets indicate the location of most 5′ nucleotide of the primer according the footnotes of Table 1 or Table 2):

λ1-434 (−434):
5′-GGCAATGATTCTACCTTGTGTAG-3′;
λ1-404 (−404):
5′-AGATGTGGATAGATACTGATGAC-3′;
λ1-365 (−365):
5′-AACTGCTCTCTCCTGAAGTGC-3′;
λ1-322 (−322):
5′-TTTCCAGCATGAGCTGCAGAG-3′;
λ1-253 (−253):
5′-GTCTCCTCAAGCTGTCACTGG-3′;
λ1-205 (−205):
5′-CCTGTGAAGATCTTCCATACC-3′;
STL1F1(−100):
5′-GTCACCACCTTCCAAGAATTACC-3′;
λ1-51 (−51):
5′-CTTGAGAATAAAATGCATGCAAGG-3′;
λ2-56 (−56):
5′-GAGAACAGGACCAGGTGCTG-3′;
λ2-140 (−140):
5′-AGGATTACCACCCACTGCTTC-3′;
λ2-280RI (−168); λ2P-START (−117); λ2P-NheI (−88) (see all below);
λ3-47 (−47): 5′-CCCAGGTGCTTGCCCCAC-3′;
λ3-89 (−89):
5′-GCTAGCTAGTAGTTTGATTCATGT-3′;
λ3-126 (−126):
5′-ACCCACTTCAAGTGAGGTCAC-3′;
Jκ1-134 (−134):
5′-CTCTGTCAGAGAAGCCCAAGC-3′;
κ0.8-162 (−162):
5′-TTGAAGTCTCAACTATGAAAATCAGC-3′;

κ1.1-410 (−410):
5′-GAGGCTTAGCCCCTCTGAGGTTAG-3′;
κ1.1-453 (−453):
5′-AGAGTAAGCTGGAACCCACAACAGC-3′;
κ1.1-474 (−474):
5′-GAATTCAGTTTATCACACGTATAGAG-3′;

κ0.8 for (−49); κ1.1 for (−365), both described in (Grawunder et al., 1995a). The reverse primers were: λ1: Cλ1479; λ2: Cλ2RE2; λ3: Cλ3RE3; κ^0: Cκrev; *mb-1*: mb-1rev, mb-1rev2 (see above).

A primer set specific for the ubiquitously expressed hypoxanthine-guanine-phosphoribosyl-transferase (HPRT) gene was used as positive control (Engel et al., 1999). Conditions were selected under which the different primer pairs exhibited comparable efficiencies, as tested using appropriate plasmid DNAs as templates. The intensities of PCR bands were determined using EASY software (Herolab, Wiesloch, Germany). In each case, only one PCR product of the expected size was observed; the identities of selected products were confirmed by sequencing.

2.4. Construction of luciferase reporter vectors

The genomic EcoRI fragments carrying the Jλ and Cλ gene segments J–C λ4: 2.8 kb; Jλ3: 5.0 kb; Cλ3–Jλ1–Cλ1: 8.6 kb, all cloned in pUC12 (Weiss and Wu, 1987), were partially sequenced by primer walking using dye terminators and either an ABI 373 or 377 automatic sequencer (Applied Biosystems). Sequences of the primers can be supplied on request. Updated sequences of both λ J–C gene clusters have been submitted to EMBL/GenBank database (Accession. No: λ1/λ3: X58411; λ2/λ4: X58414). The region carrying the λ2 germ-line promoter was PCR-amplified using Pfu-polymerase (Stratagene, La Jolla), 200 μM of each dNTP and each primer pair at 250 nM concentration. PCR products were digested by *Sst*I and cloned in the *Sst*I and *Sma*I sites of the 5′ multi cloning site (MCS) of the luciferase reporter vector pGL3Basic (Promega) and of pGL3Eλ2-4. The pGL3Eλ2-4 vector was generated using a BamHI/SalI fragment containing the 3′ enhancer of the λ2-4 gene cluster (Hagman et al., 1990) from pUC18 and cloning it in the same sites of the 3′ MCS of pGL3Basic in its correct orientation. Primers were as follows (exact locations 5′ of Jλ2 can be determined from the genomic sequence described above):

λ2FOR-SACI:
5′-GCCGCCGAGCTCTGTGGCGCATTCTTTCATAAG-3′;
λ2P-50RI:
5′-GCCGCCGAGCTCTGGTTCTAAAGAAGCTGACC-3′;
λ2P-104RI:
5′-GCCGCCGAGCTCAACACTCAGCCTGTAAAATC-3′;
λ2P-152RI:
5′-GCCGCCGAGCTCAACAAAGACACCTGAATT-3′;
λ2P-203RI:
5′-GCCGCCGAGCTCTACACAGCTGTCAGCATAG-3′;
λ2P-250RI:
5′-GCCGCCGAGCTCATGGTGTTGAATCAACTCC-3′;
λ2P-265RI:
5′-GCCGCCGAGCTCCATGGGGTACCTGCCTG-3′;
λ2P-280RI:
5′-GCCGCCGAGCTCTGCCTGACAAGAAGACAAGCC-3′;
λ2P-TATAA:
5′-GCCGCCGAGCTCCTATAAAAGGATTACCACCCAC-3′;
λ2P-START:
5′-GCCGCCGAGCTCAAGTGAGGTCATAGCTC-3′;
λ2P-NheI:
5′-GCCGCCGAGCTCTAGCTAGTAGTTTGATTCAGC-3′;
used with λ2REV as reverse primer:
5′-GCCAGCACCTGGTCCTGTTC-3′;

Transfection efficiencies were determined using the e-GFP expression plasmid peGFPCMV. This plasmid is based on the peGFP vector (Clontech) which encodes e-GFP controlled by a 0.6 kb EcoRI/BamHI fragment carrying the hCMV early promoter derived from pCMVβ (Clontech).

2.5. Transient transfection and measurement of luciferase activity

Cells were grown to their optimal density for efficient transfection ($1–3 \times 10^6$ cells/ml, depending on the cell line) in 175 cm^2 tissue culture flasks for cells growing in suspension (Sarstedt, Nürnberg, Germany). Optimal conditions for transfection were determined by transiently transfecting cells with the e-GFP expression plasmid (see above) following the protocol described below. The percentage of GFP-positive cells was quantitated 24 h after transfection using either a FACScalibur flow cytometer (Becton and Dickinson, Heidelberg, Germany) or fluorescence microscopy. Before transfection cells were harvested, cell suspensions were washed in serum-free Optimem-1 medium, counted and diluted to a density of 1×10^7 cells/ml in serum free Optimem-1 and 333 μl of this suspension was transferred to a well of a 96-deep-well-plate (Ad-

vanced Biotechnologies, Hamburg, Germany). Subsequently, 167 μl of the transfection mixture containing 25 μl Tris–Cl pH 7.5, 200 ng luciferase reporter vector, 1.25 ng of the Renilla luciferase expression plasmid pRLCMV (Promega) and 140 μg DEAE-Dextran (10 mg/ml in PBS; Pharmacia, Freiburg, Germany) were added and mixed. Covered plates were incubated in a 37 °C water bath for between 20 and 35 min, depending on the cell line. Cells were washed once in serum-free Optimem-1 at room temperature, then cultured on the same plate in 1.25 ml Optimem-1 medium containing supplements. The medium was replaced once after 12 h. After 24–26 h, cells were harvested to determine the luciferase activity in cell lysates. To this end, the cells were washed once in PBS and lysed by rocking for 20 min in 50 μl passive lysis buffer supplied with the dual luciferase assay system (Promega).

Dual luciferase activity was measured by analyzing 18 μl of the lysates using a luminometer capable of reading 96-well-plates with dual auto-injection (Berthold, Bad Wildbad, Germany). As recommended, 100 μl reagent was injected at each step and luciferase activity was measured for 10 s after a delay of 2 s. Duplicate experiments were performed two to five times for each cell line. Highly reproducible results were obtained in this way.

2.6. Computer algorithms for the prediction of S/MARs

Two contig sequences were computationally analyzed for the presence of S/MARs. The first contig contained the region starting 5′ of Jλ3 and proceeding to Cλ1 and the second contained the region starting 5′ of Jλ2 and continuing to Cλ4.

Two different S/MAR-prediction computer algorithms were used for this purpose. The approach of Benham et al. predicts S/MARs based on the stress-induced duplex DNA destabilization properties that characterize these regions (Benham, 1992; Benham et al., 1997). Because base unpairing is known to occur at S/MARs, their duplex destabilization properties may be expected to reflect their activities. S/MARs exhibit a strong decrease in $G(x)$, the predicted incremental free energy needed to open the base pair at position x in the sequence under negative superhelicity. The destabilization patterns exhibited by S/MARs are characteristically long, extending over >100 base pairs.

The second method, the MarFinder program of Singh et al. (1997) predicts S/MAR locations based on the presence of seven frequently observed sequence features that are present in some, but not all, S/MARs. These are: AT-richness, an origin of replication rule, natural curvature, predicted topoisomerase II recognition sites, kinking (i.e. occurrence of the dinucleotides TG, CA or TA at defined spacings), TG-richness, the presence of AT_nA tracts for $n = 2$, 3 or 4 and the ATC rule (>20 sequential occurrences of 'H', i.e. A, T, C but no G). MarFinder slides a window over the sequence and computes these properties within each window. Here, the sliding window parameters used are a window width of 100 bp and a slide distance of 10 bp.

Because stress-induced destabilization occurs preferentially in AT-rich regions, both methods could, in principle, find similar sites. However, the principles underlying these two approaches are quite different. The sequence properties assessed in MarFinder are entirely local, while stress-induced destabilization is not a local attribute. Because imposed stresses couple together the behaviors of all base pairs that experience them, the transition properties of a site depend not just on its local properties, but also on how it competes with all other sites that experience the stress. This makes the destabilization properties of stressed regions highly context-dependent. So, these two approaches should be regarded as evaluating distinct properties of the DNA, one based on local sequence and the other on longer-range structural competitions.

2.7. Characterization of S/MAR activity in vitro

For determination of the binding potential of the DNA fragments to nuclear scaffolds in vitro, a plasmid carrying a 2.7 kb fragment, derived from the region between Cλ3 and Jλ1, was digested with *Pst*I. As control, the S/MAR element located in the large J–C intron of the κ L chain locus and some neighboring regions were isolated as a 1.1 kb PvuII/AvrII fragment from the plasmid pRBJCκ (Lewis et al., 1982). These fragments were tested for their potential to bind to nuclear scaffolds in vitro following the protocol described in (Kay and Bode, 1995). Briefly, nuclei were prepared and nuclear scaffolds were isolated by extraction. The digested DNA samples including the pTZ18R plasmid, which carries a set of S/MAR standards, were radioactively labeled using T4 polymerase. Aliquots of the respective sample DNA and the standard were mixed and incubated with the nuclear scaffolds in the presence of a vast molecular excess of unlabeled competitor DNA that leads to a saturation of non-specific binding sites without the loss of the specific binding (Kay and Bode, 1995). Experiments were performed in the presence or absence of EDTA, since binding specificity is sometimes modified due to the chelating properties of EDTA. Nuclear scaffolds were pelleted and samples were separated on agarose gels. After electrophoresis, gels were dried, subjected to autoradiography and analyzed using a phophorimager. Estimation of binding strength was carried out by comparing association of the radioactively labeled λ S/MARs to the scaffold to that of the κ intron S/MAR, which had been determined as 50% in a normalized set of standards

previously. The normalization was achieved by using several fragments (Mielke et al., 1990; Bode et al., 1992) and synthetic S/MAR as well as non-S/MAR standards with binding strength ranging from 0 to 100% according to the 'modified equal counts approach' as described by Kay and Bode (1995). A total of 50% binding strength under such conditions indicated that half of the radioactive fragment had bound to the scaffold and half remained in the supernatant.

3. Results

3.1. Mapping the initiation sites of germ-line transcripts of the murine J–C λ clusters

In order to identify the promoters and other elements that regulate germ-line transcription of the λ J–C gene clusters, it was first necessary to determine the precise initiation sites of their transcripts. Mapping of transcriptional start sites is often difficult, especially for low abundance transcripts with initiation site heterogeneity. Even when starting with large amounts of mRNA, this heterogeneity may cause the yields of specific fragments to fall below the detection levels of commonly used assays. In our case, conventional techniques, such as primer extension or ribonuclease protection assays, failed to provide conclusive results (data not shown). Similarly, the conventional 5′-RACE technique that was previously used to characterize λ J–C germ-line transcripts may select short molecules or molecules with incomplete 5′-ends.

To circumvent such problems we performed a new 5′-RACE assay using the SMART-RACE technique. This method utilizes both the inherent oligo(dC) terminal transferase activity of the reverse transcriptase during cDNA synthesis and the ability to switch templates during DNA synthesis. The first-strand cDNA synthesis is primed by an oligo(dT) primer, with a second primer (SMART II) included in the first-strand synthesis reaction. The latter primer contains a short G stretch at its 3′-end that can pair with the oligo(dC) stretch added by the reverse transcriptase. This enables the reverse transcriptase to switch templates and incorporate the SMART II primer in the cDNA. These attributes should enable the SMART procedure to produce high yields of full length cDNAs (Chenchik et al., 1998).

We applied this technique to generate a SMART-cDNA library for amplification of the 5′ regions of Ig germ-line transcripts. For this purpose, we used the total RNA of differentiating *R2-bfl* cells grown for 5 days in the absence of IL-7 (Grawunder et al., 1995b; Engel et al., 1999). First, a primer specific for the amplification of germ-line transcripts of the individual λ clusters was used and the PCR fragments were cloned and sequenced. The 5′-most nucleotide of each clone that showed identity with the particular DNA sequence, but was not found in the primer, was taken as the transcriptional start position.

The 5′-RACE-PCR procedure applied to J–C λ1 using three different Cλ1-specific primers produced two PCR products lengths which differed by ≈250 bp. Clones of the smaller fragment consisted of J–C λ1 splice products that initiated either a few nucleotides upstream of Jλ1 or within Jλ1 (Table 1). Due to the redundancy of such clones, we find it highly probable that one initiation region of the J–C λ1 germ-line transcript is located just upstream of Jλ1.

The clones containing the largest inserts derived from the second fragment of the λ1 transcripts were found to include a J–C λ1 splice product that started almost 300 nucleotides upstream from the Jλ1 region (Fig. 1A, Table 1). These results indicate that a second transcriptional initiation site exists for germ-line λ1 transcripts.

Multiple start sites were also identified for both the J–C λ2 and the J–C λ3 transcripts (Fig. 1B, Table 1). In both cases most clones started ≈50 bp upstream from the respective Jλ. The 5′-most initiation site of the λ2 germ-line transcript was located at position −116. The 5′-most initiation site of a λ3 germ-line transcript mapped to position −79 (Fig. 1B, Table 1).

The start site heterogeneity observed could in principle be due to differences among individual pre-B cell clones that might comprise the *R2-bfl* line. To resolve this question, we investigated whether λ germ-line transcripts exhibit the same heterogeneity in an established cell line of clonal origin. RNA was isolated from P3X63Ag8.653 plasmacytoma cells and the transcriptional initiation sites of λ2 and λ3 germ-line transcripts were characterized using the SMART-5′-RACE technique. Initiation of λ2 and λ3 germ-line transcripts in this clonal cell line mapped to the same regions and in numerous cases, even to the same nucleotides, as were found for *R2-bfl* cells (Table 1).

3.2. Initiation sites of transcripts of the germ-line κ and H chain clusters and of mb-1 (Igα), a representative gene with heterogeneous transcriptional initiation

To assess whether our approach provides a representative view of the transcriptional start sites of germ-line transcripts, we amplified Cκ or Cμ-containing transcripts from the SMART-RACE cDNA library of *R2-bfl* cells, then cloned and sequenced them. As summarized in Table 2, the 5′-most extension of many clones of the 1.1 kb transcript mapped around the region described previously (Martin and Van Ness, 1990).

Start sites for the 0.8 kb germ-line κ transcript were previously determined by one group to be located within the KI region 5′ of Jκ1 (Martin and Van Ness, 1990) and by another to occur roughly 60 bp upstream within the

Table 1
Start sites of transcripts of Ig germ-line λ J–C clusters as defined by SMART-5′-RACE

R2-bfl				*X63/0*	
λ1^0-short Start[a]	λ1^0-long Start[a]	λ2^0 Start[a]	λ3^0 Start[a]	λ2^0 Start[a]	λ3^0 Start[a]
+19 (1)	−212 (1)	−43 (1)	+13 (3)	−50 (3)	+13 (1)
+16 (3)	−213 (1)		+4 (1)	−52 (4)	+4 (2)
−3 (1)	−216 (1)	−54 (2)	−9 (1)	−54 (2)	−6 (2)
−11 (1)	−223 (1)	−56 (5)	−12 (1)	−56 (7)	−12 (1)
−12 (1)	−224 (1)	−58 (1)			−16 (1)
−15 (1)	−225 (1)	−74 (1)	−44 (1)		−44 (1)
−17 (4)	−226 (1)	−79 (3)	−46 (1)		−46 (1)
−21 (1)	−228 (1)	−116 (1)	−52 (1)		−52 (1)
	−234 (2)				−54 (1)
	−237 (1)			−56 (2)	−56 (3)
	−238 (1)				−58 (1)
	−265 (3)			−66 (2)	−66 (3)
	−271 (2)			−69 (3)	−69 (1)
	−272 (1)			−79 (1)	
	−283 (3)				
	−293 (1)				
$n = 13$	$n = 22$	$n = 14$	$n = 17$	$n = 16$	$n = 19$

[a] Transcription start sites were numbered by defining the most 3′ nucleotide of the recombination signal heptamer sequence upstream of each Jλ gene segment as −1. The number of clones found with this initiation site is given in brackets.

Table 2
Start sites of germ-line transcripts of Ig J–C κ and μ clusters as defined by SMART-5′-RACE

κ^{0}1.1		κ^{0}0.8			μ^0		Iμ	
Start[a]	Start [a,b]	Start[c]	Start[c,d]	Start[b,c]	Start[e]	Start[e,f]	Start[g]	Start[g,h]
−260 (1)		+2 (1)			+103 (1)		−512 (1)	
−300 (1)		−3 (2)			+58 (1)		−608 (1)	
−374 (1)		−9 (1)			+6 (2)		−683 (1)	
−375 (2)		−20 (1)			−3 (1)		−689 (3)	−689 −688
−376 (5)		−24 (6)			−35 (2)		−695 (2)	−695 −692
−377 (2)		−27 (2)			−39 (4)	−38	−708 (2)	−705 −703 −702
−385 (3)		−48 (2)		−50	−43 (1)		−712 (1)	−713 −712 −711 −710
−386 (1)		−58 (1)	−70	−52	−57 (16)	−57	−717 (1)	
−389 (2)	−388	−96 (2)	−87		−63 (1)	−60	−718 (1)	−718
−391 (2)		−100 (1)	−104 ± 1		−71 (1)	−87	−731 (1)	−731 −730 −729 −727
−393 (4)		−102 (1)	−113 ± 1					
		−194 (1)						
$n = 24$		$n = 21$			$n = 30$		$n = 14$	

[a] Transcription start sites were numbered defining the first nucleotide upstream of the Cκ region as −1 and the number of clones found with this initiation site is given in brackets.

[b] Data taken from Martin and Van Ness (1990) are shown for comparison.

[c] Transcription start sites were numbered by defining the most 3′ nucleotide of the recombination signal heptamer sequence upstream of Jκ1 as −1 and the number of clones found with this initiation site is given in brackets.

[d] Data taken from Leclercq et al. (1989) are shown for comparison.

[e] Transcription start sites were indicated by defining the most 3′ nucleotide of the recombination signal heptamer sequence upstream of D_{Q52} as −1 and the number of clones found with this initiation site is given in brackets.

[f] Data taken from Kottmann et al. (1994) are shown for comparison.

[g] Transcription start sites were numbered by defining the most 3′ nucleotide of the intron-derived Iμ exon as −1 and the number of clones found with this initiation site is given in brackets.

[h] Data taken from Lennon and Perry (1985) are shown for comparison.

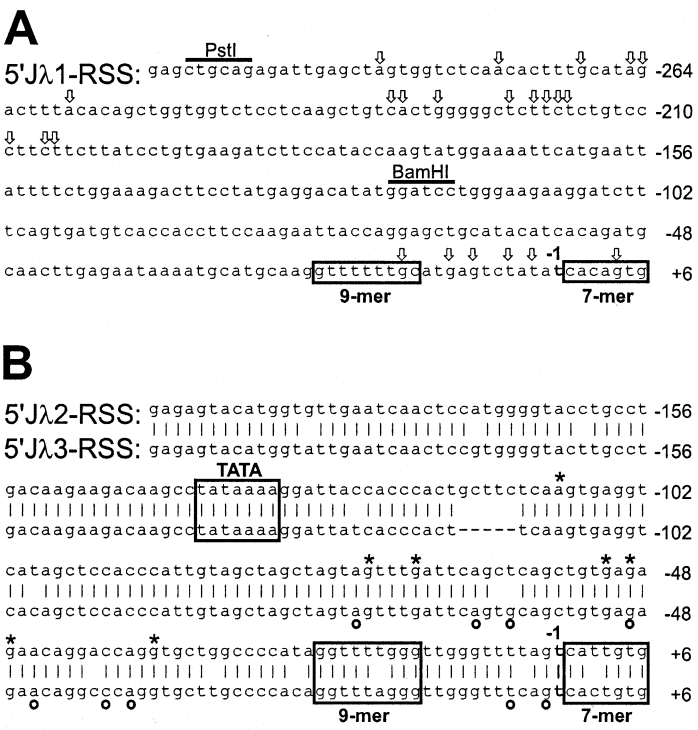

Fig. 1. Sequence of the 5′ part of J–C λ germ-line transcripts. The sequence upstream of each functional Jλ gene segment containing the recombination signal sequences (RSS) is shown. The RSS are indicated as 9- and 7-mer, respectively. Numbering starts with −1 with the first nucleotide upstream of the 7-mer RSS. (A) Partial sequence of the germ-line λ1 J–C transcript and 5′ flanking regions. *Bam*HI and *Pst*I indicate the presence of restriction sites for these enzymes. Arrows point to the 5′ end of cDNA clones isolated by SMART-5′-RACE. (B) Sequence alignment of partial sequences of λ2 and λ3 J–C germ-line transcripts and flanking regions. The TATA-box consensus sequence is shown boxed. Stars or open circles above or below the sequences indicate the 5′ end of cDNA clones produced by SMART-5′-RACE of λ2 or λ3 germ-line transcripts, respectively. Additional sequence information for both complete λ J–C cluster is available from EMBL/GenBank database (see Section 2).

KII region (Leclercq et al., 1989). We found clones starting in both of these regions (Table 2).

Multiple start sites were found for the μ^0 transcript (Table 2), with a major initiation site located at position −57. The initiation region of this transcript was found to coincide with that previously described by Kottmann et al. (1994), with some sites (such as at −38 and −57) being precisely identical. Comparison of the initiation sites found for the Iμ transcript by our SMART-5′-RACE and those described by Lennon and Perry (1985) also revealed almost complete identity (Table 2).

Characterization of the Ig germ-line transcripts by the SMART-5′-RACE technique yielded some results that did not fully agree with previously published data (Table 2) (Leclercq et al., 1989; Martin and Van Ness, 1990). In order to verify the capacity of our approach to yield a representative view of transcriptional start sites even further, we analyzed transcriptional initiation sites of the *mb-1* (Igα) gene, a gene that is distinct from the germ-line Ig gene segments and which is known to initiate transcription from multiple sites (Travis et al., 1991). Data obtained using two different specific reverse primers agreed with previously published data, although differences in their frequencies of use were observed. Designating the 5′-most major initiation site previously described (Travis et al., 1991) as position +1, the 5′-most initiation site identified by SMART-5′-RACE mapped to position +2 (2/16 clones, data not shown), which has been suggested to be one of the two major start sites in the 5′ part of the initiation region. Surprisingly, most of the clones (11/16) initiated at position +12, which was previously described as a weak transcription initiation site (Travis et al., 1991). An additional site identified by 3/16 clones was found to be located at position +16, one of the supposedly strong initiation sites in the 3′ part of the initiation region of *mb-1*. Taken together, these data confirm that the 5′-RACE technique used by us provides a representative picture of transcription initiation sites.

3.3. Few λ and κ J–C germ-line transcripts initiate upstream of the 5′-most start site mapped by the 5′-RACE

The discrepancies among the results obtained using different RACE techniques, together with the heterogeneity we observed in the transcriptional initiation

sites of the λ J–C germ-line transcripts, prompted us to test whether the sites we mapped really represent full-length transcripts. For this purpose, we designed an isotype-specific, semi-quantitative RT-PCR technique. Primer pairs were selected in which the reverse primer annealed in the respective C region and the forward primers annealed either within the defined transcripts or at increasing distances upstream from the 5′-most start site, as found above (for the exact position of the primers see Section 2). As shown in Fig. 2(A), analysis of the J–C λ1 PCR amplification products revealed transcripts that obviously extend beyond these 5′-most transcription start site. However, yields of amplification products using forward primers located upstream from the 5′-most initiation site were much lower than the yields achieved when primers located within the predicted cDNA sequence were used (Fig. 2(A); see difference between signal for −253 and −365). Similar results were obtained for the J–C λ2 and J–C λ3 germ-line transcripts (Fig. 2(B,C); see difference between −117 and −143 for λ2 or −47 and −89 for λ3, respectively). We therefore conclude that only a small fraction (∼5%) of the J–C λ germ-line mRNA initiates upstream from the 5′-most sites identified by the SMART-5′-RACE technique.

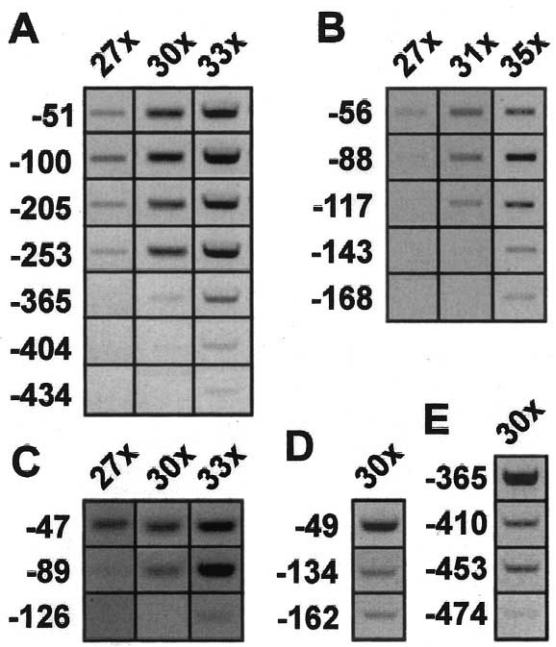

Fig. 2. Determination of the major initiation regions of Igλ germ-line transcripts by RT-PCR. Germ-line transcripts of the J–C clusters of κ and λ were amplified with C region specific reverse primers in combination with different primers located in or upstream of the 5′-most extended cDNA sequence described in Table 1. All PCRs showed comparable efficiencies at the selected conditions, were carried out in duplicates (not shown) and repeated at least once with different cDNA preparations. As template, cDNA was generated from *R2-bfl* cells grown 5 days without IL-7 (shown) or from total BALB/c spleen cells (data not shown). Both preparations yielded comparable results with the λ1-322 primer (see below). (A) Few J–C λ1 germ-line transcripts initiate upstream of the 5′-most initiation site defined by SMART-RACE. No dramatic differences were observed using primers located within the defined sequence (−51; −100; −205; −253, numbers indicate the distance in respect to the 7-mer RSS according to the footnote of Table 1). An at least 16- to 20-fold reduction was observed using primers located upstream of the most 5′ start site (−365; −404; −434). One additional primer (−322) only yielded a PCR product using BALB/c spleen cDNA, which might be due to a sequence difference in the sequence complementary to this primer between BALB/c and *R2-bfl* cells (data not shown). (B) Few J–C λ2 germ-line transcripts initiate upstream of the 5′-most start site defined by SMART-RACE. Comparable intensities were obtained using the −56 or −88 primers located in the major initiation region. A slight reduction was found with the −117 primer, located at the most 5′ initiation site, while a dramatic drop in signal intensity was found with the −143 and −168 primer that were located upstream of the cDNA clone with the largest 5′ extension. (C) Few J–C λ3 germ-line transcripts initiate upstream of the 5′-most start site defined by SMART-RACE. Strongly reduced amounts of PCR amplification product were obtained using primers located upstream of the 5′-most initiation region (see −89 and −126 primer). (D) Some κ0.8 kb and (E) κ1.1 kb germ-line transcripts initiate upstream of the 5′-most initiation site.

Comparable data were obtained when the same approach was applied to the 0.8 kb (Fig. 2(D) and data not shown, difference between −49 and −134 primer) and the 1.1 kb (Fig. 2(E) and data not shown; see difference between −365 and −410) germ-line transcripts of the κ locus.

3.4. Promoter activity at the 5′ region of Jλ2

As a first step in determining the regulatory elements governing germ-line transcription from the λ locus, we studied the potential of the region located upstream of Jλ2 to drive transcription of a reporter gene in cells of the B lineage. We chose this region for three reasons: first, its homology to the region upstream of Jλ3, second only moderate heterogeneity among the λ2 germ-line transcriptional start sites, and third, most importantly, there are no start/stop codons in this region. The potential promoter region located upstream of Jλ1 contains multiple start/stop codons, which might be the reason why we could not demonstrate any promoter activity in our reporter assay (data not shown).

Promoter activity of the Jλ2 germ-line region was tested by transiently transfecting cell lines representing different stages of B cell differentiation using the constructs based on the pGL3 luciferase reporter vector shown in Fig. 3(A). This plasmid was used due to its high luciferase expression level and low background (Groskreutz et al., 1995). Enzymatic activity of luciferase was determined in cell lysates 24 h after transfection. Promoter activity was tested in the region 5′ of Jλ2 between the transcriptional initiation site of Jλ2 and the EcoRI site (i.e. between position −35 and position −446, using the nomenclature described in the footnote of Table 1). Promoter activity of this DNA fragment was determined either alone or combined with the λ2-4 enhancer inserted downstream of the luciferase gene (Fig.

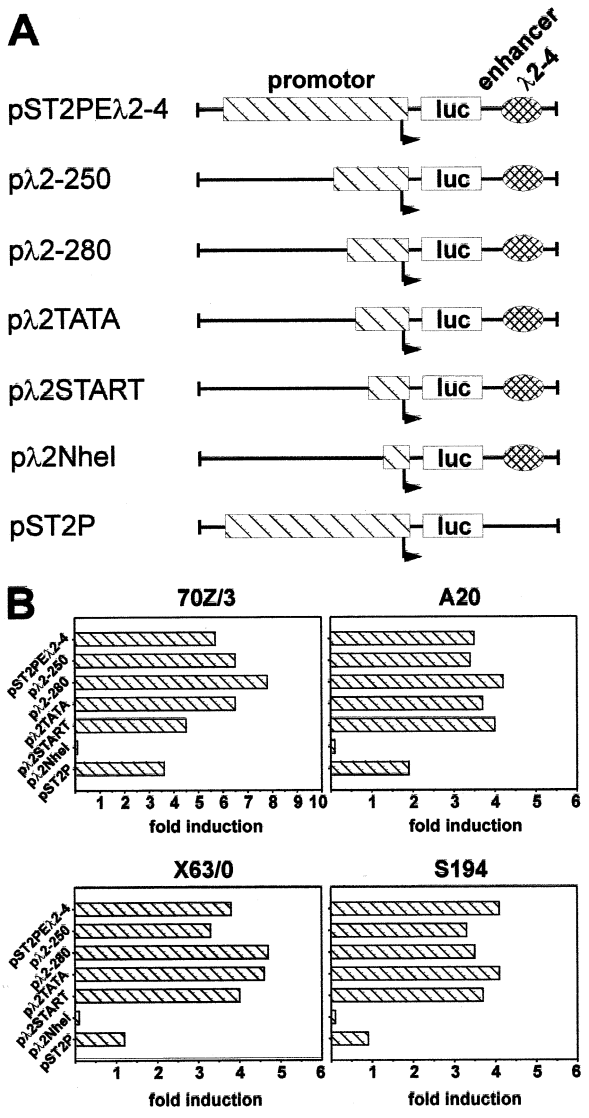

Fig. 3. Promoter activity upstream of Jλ2. (A) Schematic representation of the pGL3 based luciferase expression constructs and variants carrying sequential 5′ deletions in the potential promoter region. Additional intermediate variants between the upper construct containing a 0.4 kb fragment and the second construct were also generated and tested (data not shown). (B) Promoter activity of luciferase expression constructs controlled by different variants of the λ2 germline promoter region. Activity was determined using different cell lines representing different stages of B cell development (*70Z/3*: preB; A20: mature B; *X63/0*, S194: plasma cell). Normalized values were obtained by cotransfection of a renilla luciferase encoding expression plasmid. The induction factor of luciferase activity compared to the pGL3 vector carrying the λ2-4 enhancer or, for pST2P, the empty pGL3 plasmid, is shown. Representative data from one experiment carried out in duplicates are shown. Experiments were performed two to five times, showing highly reproducible results.

3A). Significant promoter activity was detectable using either the complete fragment or several subfragments obtained by sequential deletions from the 5′ end of the potential promoter region (Fig. 3B). The observed activity was rather weak compared to that of the hCMV promoter (from 23- to 2000-fold lower, depending on the cell line used). Transcription from the Jλ2 germ-line promoter appears not to depend on the presence of the TATA motif found in the promoter region (Fig. 1A), as no change in promoter activity was observed when this site was deleted (Fig. 3(B), pλ2START). However, a dramatic drop of activity was observed upon deletion of the region spanning from −120 to the NheI site at −84. The activity measured in this case did not exceed the background level of a reporter plasmid that contains only the λ2-4 enhancer without any promoter region. Thus, measurable promoter activity of germ-line J–C λ2 is contained within the fragment ranging from −120 to −35.

Promoter activity of the full-length λ2 germ-line promoter construct could also be observed under these conditions in the thymoma cell line *BW 5147* and the pre-B cell line *70Z/3*, both of which do not express endogenous germ-line transcripts of λ (Fig. 3(B) and data not shown). However, this activity was less dependent on the presence of the λ2-4 enhancer.

3.5. Transcriptional initiation of the λ1 germ-line transcript downstream of a potential nuclear scaffold/matrix attachment region

During the process of defining the regulatory elements and transcription factor binding sites that occur upstream of the initiation regions of the J–C λ germ-line transcripts, DNA sequences were determined in both λJ–C gene clusters. The regions between Jλ1 and Cλ3 and between Cλ2 and the Jλ4 pseudo-gene were found to contain long sites displaying extreme A/T-richness, an attribute suggesting susceptibility to stress-induced destabilization which has been associated with S/MARs, the sites where chromosomal DNA is attached to the nuclear matrix or scaffold. S/MARs are thought to have important modulating effects on transcriptional regulation of the eukaryotic genome.

Two different computational methods have been developed to predict sites with S/MAR potential from the obtained DNA sequences (Benham et al., 1997; Singh et al., 1997). The approach of Benham et al. predicts S/MARs based on the stress-induced duplex DNA destabilization properties that characterize these regions (Benham, 1992; Benham et al., 1997) while the second method, the MarFinder program of Singh et al. (1997), predicts S/MAR locations based on the presence of seven frequently observed sequence features that are present in S/MARs.

Both methods were applied to the two λ contigs (λ2/λ4 and λ3/λ1) and are shown in Fig. 4(A,B). Despite the differences between these methods, both yielded comparable results. Potential S/MARs were identified between Cλ2 and Jλ4 (Fig. 4A) and Cλ3 and Jλ1 (Fig. 4B). The fact that both methods flag the same location may

be taken as evidence that a S/MAR exists at that position.

To experimentally test for the presence of a S/MAR element between the λJ–C gene clusters, a plasmid carrying the region from Cλ3 to Jλ1 was digested with *Pst*I. This fragment was tested for its potential to bind to nuclear scaffolds in vitro following the protocol described in (Kay and Bode, 1995). In order to rigorously evaluate the binding strengths of this segment, we compared it to a known standard, the S/MAR element located in the large J–C intron of the κL chain locus (Cockerill et al., 1987), that was isolated as a 1.1 kb fragment from the plasmid pRBJCκ (Lewis et al., 1982). This well-characterized S/MAR, which has 50% binding strength in a set of normalized standards, was included in the present experimental evaluation for comparison. Fig. 4(C) shows that the λ-associated fragment displayed activities by far exceeding that of the well-established κ-S/MAR element, suggesting they do in fact contain S/MARs.

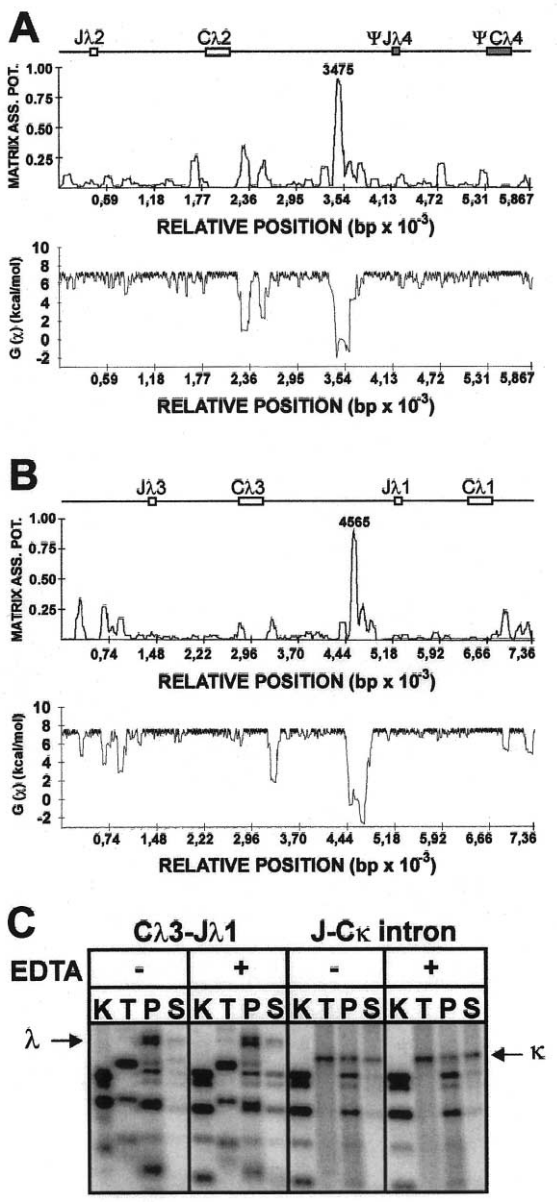

Fig. 4.

4. Discussion

Despite the almost identical recombination motifs that occur in the various loci and gene segments encoding immunoglobulin or T cell receptor chains, the rearrangements at a given locus are largely restricted to particular developmental stages of the respective cell lineage. Knowledge of the molecular events involved in the sequential and stage-specific recombination of such gene segments is still very limited. Therefore, it is essential to characterize the elements that are involved in the regulation of these processes.

Our goal is to characterize the activation of rearrangement at the murine λ L chain locus. Since transcription of the gene segments was found to be important for efficient rearrangement, identification of the promoters that drive this transcription should illuminate the regulatory processes involved. However, before we could determine these promoter regions, we first had to characterize the initiation sites of the λ J–C germ-line transcripts. Using the SMART-5′-RACE technique we extended our previous preliminary results, showing that transcripts from the germ-line λ2 and λ3

Fig. 4. Analysis for the presence of S/MAR elements in the Igλ L chain cluster. (A) Prediction of S/MARs in the Jλ2–Cλ4 contig based on stress induced duplex DNA destabilization (upper panel, Benham et al., 1997) or based on the application of a set of AND-OR formulations using sequence motif pattern common to S/MARs (lower panel; Singh et al., 1997). (B) Same analysis as in (A) for the Jλ3–Cλ1 contig (lower panel). Both methods identified a S/MAR in the region between Cλ3 and Jλ1 as well as Cλ2 and Jλ4. The numbers above the upper and lower panel give the position of maximum attachment to the nuclear matrix. (C) Binding of DNA fragments to nuclear scaffolds in vitro. A 1.7 kb PstI fragment derived from the genomic region located between Cλ3 and Jλ1 (left part) and, as control, the 1.1 kb fragment carrying the S/MAR located in the large J–C intron of the κ L chain locus (right part) were tested for their capacity to bind to nuclear scaffolds in vitro. The respective bands are marked by arrows. The binding strength of the κ intron S/MAR in our system was determined as 50% (Kay and Bode, 1995). Compared to this, the strength of the S/MAR identified in the λ locus was found to be much stronger. For each sample, four lanes are shown with or without the addition of EDTA. (K) S/MAR standards; (T) total input sample DNA; (P) a mixture of the total sample DNA and the S/MAR standard control DNA bound and pelleted together with the nuclear scaffolds; (S) fraction of this mixture of DNA remaining in the supernatant after centrifugation. The bands in the P and S lanes should sum up to the sum of K and T. Note that labeling efficiency of some fragments was lower than for others and might be difficult to detect at the exposure displayed in this figure. Fragment size of standard: 800, 600, 480, 320 and 160 bp; λ S/MAR: 1.7 kb; κ S/MAR: 1.1 kb.

J–C cluster start a short distance upstream from the particular J and display moderate heterogeneity at their initiation sites (Engel et al., 1999). Germ-line λ1 J–C transcripts start upstream of Jλ1 at two alternative regions roughly 200 bp apart, with considerable heterogeneity.

The validity of these results was confirmed in two ways. First, the initiation sites of previously characterized germ-line transcripts from Igμ and κ gene segments were analyzed using the same cDNA and the same RACE technique. The data obtained were in close agreement with the previously published data. Similarly, initiation sites of the characterized non-Ig gene *mb1* (Igα) were analyzed and found also to agree with published results, although the frequencies of initiation from the major sites seen in our analysis were slightly different. Second, RT-PCR was performed using primers that hybridize either downstream or upstream from start sites defined by us. These experiments confirmed that the major transcripts of germ-line λ, 1.1 kb κ and 0.8 kb κ initiate at the identified regions. Only minor signals were obtained for longer transcripts. Thus, longer transcripts are present but in low frequency, at least for transcripts of the Ig L chain germ-line loci.

Based on these findings, we identified a region upstream from the initiation site of J–C λ2 that displayed promoter activity in a reporter assay. The level of promoter activity could be enhanced the presence of the λ2-4 enhancer that was used in this system. A sequence comparison between this promoter region of λ2 and the analogous region of λ3 revealed extensive homology. This suggests that the regulation of germ-line transcription at both clusters should be similar.

We also attempted to define the promoter activity at the J–C λ1 cluster by testing the fragment of DNA upstream from Jλ1 for promoter activity using our luciferase assay. However, we did not obtain levels of reporter gene expression that exceeded the background (data not shown). This is probably due to the strong heterogeneity of the transcription initiation sites and the presence of multiple translational start and stop codons in this region.

A TATA consensus motif was observed in the J–C λ2 promoter region and a similar site was found upstream of J–C λ3. However, this motif probably is not functional since its deletion did not reduce reporter gene expression. Moreover, if the TATA motif was functional, one would expect homogeneous transcriptional initiation. A similar TATA motif is observed upstream from the major initiation site of the 0.8 kb κ germ-line transcript, which also may not be functional (Leclercq et al., 1989). No further sequence homologies have been observed so far between the promoter regions of both L chain loci. Apparently, κ and λ germ-line transcriptions are regulated differently. The appearance of germ-line transcripts of κ and λ at different developmental stages in early B cell development—κ first, then λ—might reflect these regulatory differences (Engel et al., 1999).

The 0.8 kb κ germ-line transcripts initiates within a region containing two palindromic motifs, called KI and KII. These were described previously as binding sites for KLP, a factor present in B cells (Weaver and Baltimore, 1987) that was found to be identical to Pax-5. The decisive role of this transcription factor, both in early B cell development and in the commitment of early progenitor cells to the B cell lineage, could be shown either in Pax-5 deficient mice (Morrison et al., 1998), or in Pax-5 deficient cells isolated from those mice (Nutt et al., 1999; Rolink et al., 1999). In addition, although elimination of the KI and KII sites reduced the rearrangement frequency of the targeted allele by 80%, the 0.8 kb κ germ-line transcript from this allele remained detectable in pre B cells (Ferradini et al., 1996), which might suggest that additional factors are involved in the regulation of these transcripts.

The sequences upstream from the J–C λ clusters, unlike those of the 0.8 kb κ germ-line promoter, do not contain any predicted binding sites for B cell-specific transcription factors, as determined by analysis using the TRANSFAC transcription factor database (Wingender et al., 2000). This accords with the finding that reporter constructs were active in *BW 5147* and *70Z/3* cells, although expression of endogenous J–C λ germ-line transcripts could not be detected by RT-PCR in these cells (Engel et al., 1999 and H. Engel, unpublished results). Since this activity was seemingly less dependent on the presence of the λ2-4 enhancer, it may reflect the basal activity of the core promoter region. This suggests that the tissue specificity of the J–C λ germ-line promoter might either be exclusively provided by the λ enhancers or, more likely, by regions outside of the 0.4 kb fragment examined by us.

The presence of a S/MAR between the two J–C λ cluster of both hemi-loci is consistent with this idea. S/MARs have previously been identified within the Ig loci (Cockerill and Garrard, 1986; Cockerill et al., 1987; Cockerill, 1990; Forrester et al., 1994). A S/MAR element located in the J–C intron of κ has been shown to play a role in its rearrangement (Xu et al., 1996) and a S/MAR in the J–C intron of μ is known to influence the accessibility of the J–C μ cluster at the heavy chain locus (Jenuwein et al., 1997; Oancea et al., 1997). Although no S/MAR has been found in the J–C introns of the λ clusters, the S/MAR detected upstream of λ1 and λ4 might be involved either in the long distance interaction during activation and/or in the rearrangement process at the λ locus.

The in vitro analysis of Ig promoters and of their interactions with other regulatory elements, such as transcriptional enhancers, can only suggest how these elements might interact in vivo. A conclusive determination of the roles of these elements in locus activation

and in altering the accessibility of the locus for rearrangement will require replacement or modification of these elements in the germ-line of the mouse. These studies are now in progress.

Acknowledgements

We thank Susanne zur Lage and Regina Lesch for expert technical assistance and Maren Scharfe, Petra Hagendorff and Rainer Munder for excellent help with DNA sequencing. We also want to thank Dr Antonius Rolink for the supply of *R2-bfl* cells. Many thanks are due to our colleague Steve Krawetz (Wayne State University, Detroit) for his continuous guidance while using the MarFinder algorithm. This study was supported in part by grants to S.W. from the Deutsche Forschungsgemeinschaft, the Bundesministerium für Bildung und Forschung and the niedersächsische Ministerium für Kultur und Wissenschaft. The contribution of C.J.B. was supported in part by grants RO1-HG01973 from the National Institutes of Health, and DBI 99-04549 from the National Science Foundation. J.B. was supported by a grant from Deutsche Forschungsgemeinschaft (Bo419/6-2).

References

Alessandrini, A., Desiderio, S.V., 1991. Coordination of immunoglobulin DJH transcription and D-to-JH rearrangement by promoter-enhancer approximation. Mol. Cell. Biol. 11, 2096–2107.

Bain, G., Maandag, E.C., Izon, D.J., Amsen, D., Kruisbeek, A.M., Weintraub, B.C., Krop, I., Schlissel, M.S., Feeney, A.J., van Roon, M., 1994. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell 79, 885–892.

Benham, C.J., 1992. Energetics of the strand separation transition in superhelical DNA. J. Mol. Biol. 225, 835–847.

Benham, C., Kohwi-Shigematsu, T., Bode, J., 1997. Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions. J. Mol. Biol. 274, 181–196.

Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Mielke, C., Kohwi-Shigematsu, T., 1992. Biological significance of unwinding capability of nuclear matrix-associating DNAs. Science 255, 195–197.

Chenchik, A., Zhu, Y., Diatchenko, L., Li, R., Hill, J., Siebert, P., 1998. Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR. In: Siebert, P., Larrick, J. (Eds.), RT-PCR Methods for Gene Cloning and Analysis. BioTechniques Books, Massachusetts, pp. 305–319.

Cocea, L., De Smet, A., Saghatchian, M., Fillatreau, S., Ferradini, L., Schurmans, S., Weill, J.C., Reynaud, C.A., 1999. A targeted deletion of a region upstream from the Jκ cluster impairs κ chain rearrangement in cis in mice and in the 103/bcl2 cell line. J. Exp. Med. 189, 1443–1450.

Cockerill, P.N., 1990. Nuclear matrix attachment occurs in several regions of the IgH locus. Nucleic Acids Res. 18, 2643–2648.

Cockerill, P.N., Garrard, W.T., 1986. Chromosomal loop anchorage of the κ immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell 44, 273–282.

Cockerill, P.N., Yuen, M.H., Garrard, W.T., 1987. The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. J. Biol. Chem. 262, 5394–5397.

Engel, H., Rolink, A., Weiss, S., 1999. B cells are programmed to activate κ and λ for rearrangement at consecutive developmental stages. Eur. J. Immunol. 29, 2167–2176.

Ferradini, L., Gu, H., De, S.A., Rajewsky, K., Reynaud, C.A., Weill, J.C., 1996. Rearrangement-enhancing element upstream of the mouse immunoglobulin κ chain J cluster. Science 271, 1416–1420.

Forrester, W.C., van Genderen, C., Jenuwein, T., Grosschedl, R., 1994. Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions. Science 265, 1221–1225.

Francés, V., Pandrau-Garcia, D., Guret, C., Ho, S., Wang, Z., Duvert, V., Saeland, S., Martinez-Valdez, H., 1994. A surrogate 15 kDa JCκ protein is expressed in combination with mu heavy chain by human B cell precursors. EMBO J. 13, 5937–5943.

Gorman, J.R., Alt, F.W., 1998. Regulation of immunoglobulin light chain isotype expression. Adv. Immunol. 69, 113–181.

Grawunder, U., Leu, T.M., Schatz, D.G., Werner, A., Rolink, A.G., Melchers, F., Winkler, T.H., 1995a. Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement. Immunity 3, 601–608.

Grawunder, U., Rolink, A., Melchers, F., 1995b. Induction of sterile transcription from the kappa L chain gene locus in V(D)J recombinase-deficient progenitor B cells. Int. Immunol. 7, 1915–1925.

Groskreutz, D.J., Sherf, B.A., Wood, K.V., Schenborn, E.T., 1995. Increased expression and convenience with the new pGL3 luciferase reporter vectors. Promega Notes Mag. 50, 2.

Hagman, J., Rudin, C.M., Haasch, D., Chaplin, D., Storb, U., 1990. A novel enhancer in the immunoglobulin λ locus is duplicated and functionally independent of NFκ B. Genes Dev. 4, 978–992.

Horibata, K., Harris, A.W., 1970. Mouse myelomas and lymphomas in culture. Exp. Cell Res. 60, 61–77.

Jenuwein, T., Forrester, W.C., Fernandez-Herrero, L.A., Laible, G., Dull, M., Grosschedl, R., 1997. Extension of chromatin accessibility by nuclear matrix attachment regions. Nature 385, 269–272.

Kay, V., Bode, J., 1995. Detection of scaffold-attached regions (SARs) by in vitro techniques: Activities of these elements in vivo. Methods Mol. Cell. Biol. 5, 186–194.

Kearney, J.F., Radbruch, A., Liesegang, B., Rajewsky, K., 1979. A new mouse myeloma cell line that has lost immunoglobulin expression but permits the construction of antibody-secreting hybrid cell lines. J. Immunol. 123, 1548–1550.

Kim, K.J., Kanellopoulos-Langevin, C., Merwin, R.M., Sachs, D.H., Asofsky, R., 1979. Establishment and characterization of BALB/c lymphoma lines with B cell properties. J. Immunol. 122, 549–554.

Kottmann, A.H., Zevnik, B., Welte, M., Nielsen, P.J., Kohler, G., 1994. A second promoter and enhancer element within the immunoglobulin heavy chain locus. Eur. J. Immunol. 24, 817–821.

Leclercq, L., Butkeraitis, P., Reth, M., 1989. A novel germ-line Jκ transcript starting immediately upstream of Jκ1. Nucleic Acids Res. 17, 6809–6819.

Lennon, G.G., Perry, R.P., 1985. C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5′-nontranslatable exon. Nature 318, 475–478.

Lennon, G.G., Perry, R.P., 1990. The temporal order of appearance of transcripts from unrearranged and rearranged Ig genes in murine fetal liver. J. Immunol. 144, 1983–1987.

Lewis, S., Rosenberg, N., Alt, F., Baltimore, D., 1982. Continuing kappa-gene rearrangement in a cell line transformed by Abelson murine leukemia virus. Cell 30, 807–816.

Liu, X., van Ness, B., 1999. Gene targeting of the KI–KII sequence elements in a model pre-B cell line: effects on germline transcription and rearrangement of the κ locus. Mol. Immunol. 36, 461–469.

Martin, D.J., Van Ness, B.G., 1990. Initiation and processing of two κ immunoglobulin germ line transcripts in mouse B cells. Mol. Cell. Biol. 10, 1950–1958.

Mielke, C., Kohwi, Y., Kohwi-Shigematsu, T., Bode, J., 1990. Hierarchical binding of DNA fragments derived from scaffold-attached regions: correlation of properties in vitro and function in vivo. Biochemistry 29, 7475–7485.

Morrison, A.M., Nutt, S.L., Thθvenin, C., Rolink, A., Busslinger, M., 1998. Loss- and gain-of-function mutations reveal an important role of BSAP (Pax-5) at the start and end of B cell differentiation. Semin. Immunol. 10, 133–142.

Nutt, S.L., Heavey, B., Rolink, A.G., Busslinger, M., 1999. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature 401, 556–562.

Oancea, A.E., Berru, M., Shulman, M.J., 1997. Expression of the (recombinant) endogenous immunoglobulin heavy-chain locus requires the intronic matrix attachment regions. Mol. Cell Biol. 17, 2658–2668.

Paige, C.J., Kincade, P.W., Ralph, P., 1978. Murine B cell leukemia line with inducible surface immunoglobulin expression. J. Immunol. 121, 641–647.

Prabhu, A., O'Brien, D.P., Weisner, G.L., Fulton, R., van Ness, B., 1996. Octamer independent activation of transcription from the κ immunoglobulin germline promoter. Nucleic Acids Res. 24, 4805–4811.

Ralph, P., 1973. Retention of lymphocyte characteristics by myelomas and theta + - lymphomas: sensitivity to cortisol and phytohemagglutinin. J. Immunol. 110, 1470–1475.

Rolink, A., Melchers, F., Andersson, J., 1996. The SCID but not the RAG-2 gene product is required for Sμ-Sε heavy chain class switching. Immunity 5, 319–330.

Rolink, A.G., Nutt, S.L., Melchers, F., Busslinger, M., 1999. Long-term in vivo reconstitution of T-cell development by Pax5-deficient B-cell progenitors. Nature 401, 603–606.

Schlissel, M.S., Baltimore, D., 1989. Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. Cell 58, 1001–1007.

Singh, G.B., Kramer, J.A., Krawetz, S.A., 1997. Mathematical model to predict regions of chromatin attachment to the nuclear matrix. Nucleic Acids Res. 25, 1419–1425.

Su, L.K., Kadesch, T., 1990. The immunoglobulin heavy-chain enhancer functions as the promoter for Iμ sterile transcription. Mol. Cell. Biol. 10, 2619–2624.

Travis, A., Hagman, J., Grosschedl, R., 1991. Heterogeneously initiated transcription from the pre-B- and B-cell-specific mb-1 promoter: analysis of the requirement for upstream factor-binding sites and initiation site sequences. Mol. Cell. Biol. 11, 5756–5766.

Van Ness, B.G., Weigert, M., Coleclough, C., Mather, E.L., Kelley, D.E., Perry, R.P., 1981. Transcription of the unrearranged mouse Cκ locus: Sequence of the initiation region and comparison of activity with a rearranged Vκ-Cκ gene. Cell 27, 593–602.

Villey, I., Caillol, D., Selz, F., Ferrier, P., de Villartay, J.P., 1996. Defect in rearrangement of the most 5′ TCR-J alpha following targeted deletion of T early alpha (TEA): implications for TCR alpha locus accessibility. Immunity 5, 331–342.

Weaver, D., Baltimore, D., 1987. B lymphocyte-specific protein binding near an immunoglobulin kappa-chain gene J segment. Proc. Natl. Acad. Sci. USA 84, 1516–1520.

Weigert, M.G., Cesari, I.M., Yonkovich, S.J., Cohn, M., 1970. Variability in the lambda light chain sequences of mouse antibody. Nature 228, 1045–1047.

Weiss, S., Wu, G.E., 1987. Somatic point mutations in unrearranged immunoglobulin gene segments encoding the variable region of lambda light chains. EMBO J. 6, 927–932.

Whitehurst, C.E., Chattopadhyay, S., Chen, J., 1999. Control of V(D)J recombinational accessibility of the Dβ1 gene segment at the TCR β locus by a germline promoter. Immunity 10, 313–322.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I., Schacherer, F., 2000. TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. 28, 316–319.

Xu, Y., Davidson, L., Alt, F.W., Baltimore, D., 1996. Deletion of the Igκ light chain intronic enhancer/matrix attachment region impairs but does not abolish VκJκ rearrangement. Immunity 4, 377–385.